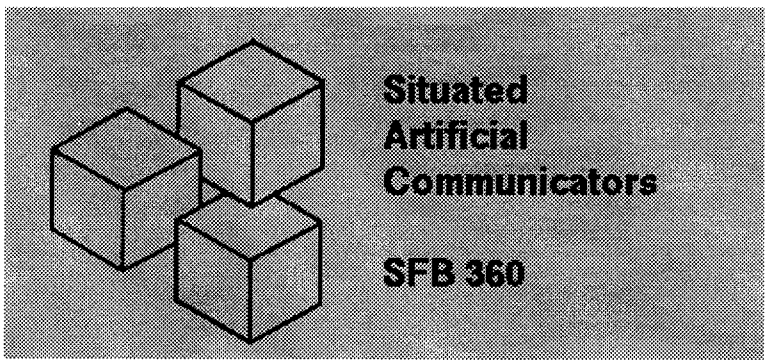
*Artificial Intelligence Review* **10:** 165–170, 1996.



# Collaborative Research Centre "Situated Artificial Communicators" at the University of Bielefeld, Germany

G. RICKHEIT & I. WACHSMUTH
*SFB 360, University of Bielefeld*
*D-33501 Bielefeld, Germany*
*E-mail: rickheit@nov1.lili.uni-bielefeld.de*
*ipke@techfak.uni-bielefeld.de*




**Abstract.** The Collaborative Research Centre "Situated Artificial Communicators" (SFB 360) is employed in modelling that which a person performs when, with a partner, he cooperatively solves a simple assembly task in a certain situation. Acoustic perception of the spoken word, visual perception of the partner and the objects and processes involved in the situation, understanding of that perceived, the formulation of own utterances, e.g. instructions to the partner, and the planning and performance of actions belong to these intelligence abilities. The theme of the Collaborative Research Centre SFB 360, in which Linguistics, Cognitive Science and Artificial Intelligence are closely entwined, is unique in this form. The scientific headquarters at the University of Bielefeld, in North Rhine Westphalia, Germany, has succeeded in entering a field of research which, in the trend of the growing importance of intelligent information systems, is a decisive factor for technical innovation.

**Key words:** man-machine communication, artificial intelligence, situatedness, cognitive robotics, speech and vision integration

## 1. Research and Method

The Collaborative Research Centre "Situated Artificial Communicators" started in the middle of 1993, after approval by the Deutsche Forschungsgemeinschaft. In our research, we examine, on the one hand, which characteristics and abilities of man, referred to here as *natural communicator*, justify his behaviour in solving an assembly task with a partner being described as *intelligent*. On the other hand, we attempt to transmit and implement the insight gained on the principles of intelligent behaviour for the construction of artificial systems, i.e. computer programmes or robots. Our long-term aim is that such *artificial communicators* take on the role of a human partner in the accomplishment of assembly tasks. Short-term, artificial communicators allow a close examination of the intelligence abilities of man. We are not conscious of several abilities which we demonstrate in every-day life as they come automatically to us. It is only when we try to replicate these artificially that we realise which intelligent abilities are necessary for the accomplishment of an assembly task.

In an assembly situation, the following intelligent abilities of man are necessary, i.e. that he (or she)

- acoustically perceive his partner, i.e. *hears*,
- visually perceive the situation, that is the partner, the objects and the on-going procedures and actions, i.e. *sees*,
- cognitively process that perceived, i.e. *understands*,
- formulate his own utterances, i.e. *speaks*, and
- plan and carry out his own actions, i.e. *grasps*, *moves* or *mounts* objects.

The fact that all these abilities are produced in a certain situation, i.e. that man is thereby *situated*, drastically limits his fundamental possibilities: the spoken word can only have certain meanings, as it refers to the actual situation; only certain own utterances are possible, as they must refer to both the task to be solved and the situation; for this reason, only certain actions are possible. In this way, situatedness is a prerequisite for a more exact examination of the intelligence abilities of man as well as for the transmission of the insights gained to artificial systems. In the foreseeable future, we will have to depart from the vision of the universally employable robot and aim at specialised robots for limited tasks.

The remarkable thing about the abovementioned intelligence abilities of man is that he can produce these even when the information available, i.e. what he has seen or heard, is incomplete or garbled. This ability is defined as *robustness*. Many artificial systems, robots or computer programmes suffer from the very fact that they are not comparatively robust. This characteristic in man, therefore, is of great interest to our research. This characteristic originates, in part, from the fact that humans can relate information from

various sources, i.e. that which they have seen or heard, and process the whole *integrated.* Disorder or incompleteness in the spoken word can be counterbalanced by that which is seen and vice-versa. Apart from this, that which is spoken can contain directions to look more closely (*active vision*). In the same way, we use properties, which we perceive in objects, e.g. their colour, their form, their size and their position, in order to refer to these objects in speech. This is the reason, for example, why we refer to an object as *the small red screwdriver in the tool box.*

As the situatedness of natural and artificial communicators plays a decisive role in the overall concept of the Collaborative Research Centre, it was necessary to select a reference situation, which touches all projects (cf. Figure 1). This situation can be defined as follows: two communicators must cooperatively construct an *object.* One, the *instructor*, has a diagram and instructs the other, the *constructor*, in such a way that the constructor, with the help of the spoken directions, can carry out the necessary actions as unproblematically as possible. Such an object can be a *model aeroplane*, which is to be constructed from the components of a wooden building kit. In this way, it is possible to make the situation to be described sufficiently interesting, but also sufficiently controllable and practically manageable. The ability to control is necessary, because our knowledge on the abilities of man in such a situation is still very fragmentary. Controlled experiments must be carried out, as man is to serve as the model for the machine constructor. The practical manageability is necessary, because the technical means, e.g. for recognition of the spoken word or for understanding a situation based on a diagram, are still a long way short of the abilities of man. In spite of this, our research group aims at constructing concrete, but limited artificial systems in a step by step procedure. Processes must be developed which will enable cooperation between different specialised systems for individual intelligent abilities.

## 2. Themes and Projects

The Collaborative Research Centre currently comprises ten projects in four thematic fields.

In the thematic field A *Speech and Visual Perception* three projects examine how an artificial communicator can receive, process and understand acoustic and visual information via sensors, i.e. microphone and camera. Here, knowledge of the most varied form is used, which must be organised in a certain manner and brought into relation to each other (Project A1). It must further be clarified how, based on the raw data of the image, i.e. information on the individual image pixels, objects can be identified and defined by the others (Project A2). Studies are being carried out on the processing of the spoken

*Figure 1.* Assembly situation (fictitious) with human instructor and one-arm robot constructor with articulated hand, receiving speech and vision input via microphone and camera (by kind permission from SFB 360 Project D3 "System Integration")

word: what contribution words such as *so*, *yes*, *nope*, *ehm* etc., intonation and speech melody make towards comprehension in a directional dialogue (Project A3).

In the thematic field B *Perception and Reference* four projects examine how the optically available information on the one hand and the spoken information on the other are to be put into relation with each other. Only in this manner can an artificial communicator correctly identify an object on hand of a spoken formulation such as *the small red screwdriver in the tool box* (Project B3) or select an appropriate formulation, so that his partner knows, which object is meant (Project B2). Experiments on how humans refer to objects with formulations and how they link between speech and image processing in a technical sense are a prerequisite for this (Project B1). Processes of active vision or "looking more closely" are examined with the help of a computer/camera system for recording eye movements (Project B4).

In the thematic field C *Knowledge and Inference* one project examines how the knowledge, of which communicators avail, allows certain conclusions in

a given situation. Our investigations examine how the connection between a word and an object, which we wish to refer to by the word, can continually alter in an on-going assembly task. For example, a wooden disc from the assembly kit is first referred to as a disc and later as the wheel of an aeroplane, or some composite object may be named an undercarriage (Project C1).

In the thematic field D *Speech-Action Systems* two projects examine the question of according to which principles the integration of individual intelligence abilities occur. To achieve this, two complementary procedures are examined. On the one hand, for situations, which are even more limited than those already described, we try - in one step - to bridge the gap between the understanding of an utterance and the resulting action (Project D1). On the other hand, based on the systems in the other projects, it is attempted to collate the functional units and to examine the performance capacity of differents forms of the collation through computer experiments (Project D3).

Further projects, prepared to expand the thematic fields C and D in the next funding period, will take on the themes of the cognitive representation of natural-language directions, the coordination of syntax in spoken discourse, and speech-action systems which are based on multiple sensor input.

## 3. Principal Investigators, SFB 360

The Collaborative Research Centre is supported by the Deutsche Forschungsgemeinschaft. Principal investigators, as of the funding period 1993-96, are the following scientists from the University of Bielefeld, Faculty of Technology and Faculty of Linguistics and Literary Studies:
Hans-Jürgen Eikmeyer, Dafydd Gibbon, Walther Kindt, Franz Kummert, Henning Lobin, Dieter Metzing, Stefan Posch, Gert Rickheit (speaker), Hannes Rieser, Helge Ritter, Gerhard Sagerer (co-speaker), Hans Strohner, Ipke Wachsmuth.

## 4. Report Series (ISSN 0946-7572)

Reports from SFB 360 can be ordered by writing to:

Administration, SFB 360
University of Bielefeld
P.O. Box 100131
D-33501 Bielefeld, Germany

or by emailing to:

anke@SFB360.Uni-Bielefeld.DE (Anke Bodzin)

The WWW homepage for SFB 360 is

http://www.techfak.uni-bielefeld.de/sfb/

More general information, abstracts of reports, and an ftp-server address for postscript versions of reports can be accessed from this WWW page. A selection of further references is given below.

## References

G.A. Fink, N. Jungclaus, H. Ritter, and G. Sagerer. A communication framework for heterogenous distributed pattern analysis. In *Proc. International Conference on Algorithms and Architectures for Parallel Processing*, pages 881–890, Brisbane, 1995.

D. Gibbon. Empirical and semiotic foundations for prosodic analysis. In U.M. Quasthoff, editor, *Aspects of Oral Communication*, pages 441–479. De Gruyter, Berlin, 1995.

A. Maßmann and S. Posch. Mask-oriented grouping operations in a contour-based approach. In *Proc. 2nd Asian Conference on Computer Vision*, volume III, pages 58–61, Singapore, 1995.

J.-T. Milde. A hybrid control architecture for a simulated robot manipulator. In *Proceedings of the 13th IASTED International Conference on Applied Informatics*, pages 370–373, 1995.

R. Moratz, H.J. Eikmeyer, B. Hildebrandt, F. Kummert, G. Rickheit, and G. Sagerer. Integrating speech and selective visual perception using a semantic network. In *1995 AAAI Fall Symposium on Computational Models for Integrating Language and Vision*, Cambridge, MA, (to appear).

U. Naeve, G. Socher, G.A. Fink, F. Kummert, and G. Sagerer. Generation of language models using the results of image analysis. In *Proc. 4th European Conference on Speech Communication and technology*, pages 1739–1742, Madrid, Spain, 1995.

H. Strohner, L. Sichelschmidt, and I. Duwe. Focus in resolving ambigous reference. In P. Bosch and R. van der Sandt, editors, *Focus and Natural Language Processing*, pages 629–638. IBM Scientific Centre/Institute for Logic and Linguistics, Heidelberg, 1994.

B. Velichkovsky, M. Pomplun, and H. Rieser. Attention and communication: Eye-movement-based research paradigms. In W.H. Zangemeister, H.S. Stiehl, and C. Freksa, editors, *Visual Attention and Cognition*. Elsevier, Amsterdam, 1995. (in press).

I. Wachsmuth and B. Jung. Dynamic conceptualization in a mechanical-object assembly environment. *This volume 10(3-4).*